

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2002-091476

(43)Date of publication of application : 27.03.2002

(51)Int.CI. G10L 15/06
G10L 15/14

(21)Application number : 2000-276944 (71)Applicant : MITSUBISHI ELECTRIC CORP

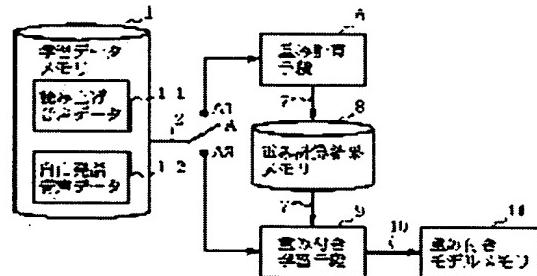
(22)Date of filing : 12.09.2000 (72)Inventor : HANAZAWA TOSHIYUKI

(54) DEVICE AND METHOD FOR LEARNING VOICE PATTERN MODEL

(57)Abstract:

PROBLEM TO BE SOLVED: To solve the problem in which an equal and robust HMM can not be obtained unless the amounts of learning data of reading voice data 100a and freely uttered voice data 100b are approximately equal for all of the phonemes in the both uttering systems.

SOLUTION: The device is provided with a learning data storage means which stores learning data for different uttering systems, a weight computing means which normalizes the reciprocals of the amounts of data for every system so that the total sum of the reciprocals becomes 1 and computes the normalized reciprocals as the weighting coefficients in accordance with the amount of data for every uttering system in the learning data and a weighted learning means which corrects the differences among the amounts of the data in the uttering systems of the learning data using the weighted coefficients computed by the weight computing means and which obtains the parameters of the voice pattern models corresponding to the learning data.



LEGAL STATUS

[Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's

[decision of rejection]

[Date of extinction of right]

Copyright (C); 1998,2003 Japan Patent Office

(19) 日本国特許庁 (JP)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開2002-91476

(P2002-91476A)

(43) 公開日 平成14年3月27日 (2002.3.27)

(51) Int.Cl.⁷

G 10 L 15/06
15/14

識別記号

F I

G 10 L 3/00

テーマコード(参考)

5 2 1 R 5 D 0 1 5
5 3 5 Z

審査請求 未請求 請求項の数 8 O.L. (全 16 頁)

(21) 出願番号

特願2000-276944(P2000-276944)

(22) 出願日

平成12年9月12日 (2000.9.12)

(71) 出願人 000006013

三菱電機株式会社

東京都千代田区丸の内二丁目2番3号

(72) 発明者 花沢 利行

東京都千代田区丸の内二丁目2番3号 三菱電機株式会社内

(74) 代理人 100066474

弁理士 田澤 博昭 (外1名)

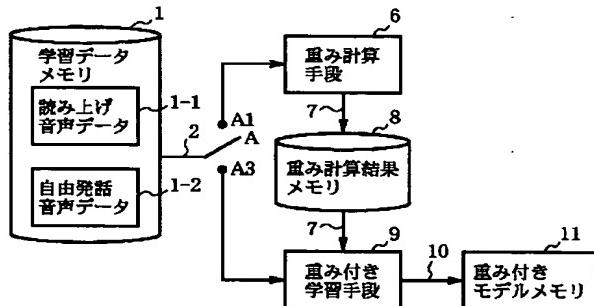
F ターム(参考) 5D015 GG00 HH23

(54) 【発明の名称】 音声パターンモデル学習装置及び音声パターンモデル学習方法

(57) 【要約】

【課題】 各音素の全てに対して読み上げ音声データ100aと自由発話音声データ100bとの学習データ量が同量程度でなければ、両方の発話様式に対して等しくロバストなHMMを得られないという課題があった。

【解決手段】 異なる発話様式の学習データを格納する学習データ記憶手段と、発話様式ごとの各データ量の逆数に、これらの総和が1となるように正規化したものを、学習データに対する発話様式ごとのデータ量に応じた重み係数として算出する重み計算手段と、この重み計算手段が算出した重み係数を用いて学習データの各発話様式間におけるデータ量の違いを補正しながら、学習データに対応する音声パターンモデルのパラメータを求める重み付き学習手段とを備えた。



【特許請求の範囲】

【請求項1】 異なる発話様式で入力された音声データの音響的特徴を表す複数種類の学習データを格納する学習データ記憶手段と、

上記複数種類の学習データにおける発話様式ごとの各データ量の逆数に、これらの総和が1となるように正規化したものを、上記学習データに対する上記発話様式ごとのデータ量に応じた重み係数として算出する重み計算手段と、

この重み計算手段が算出した上記重み係数を用いて上記学習データの上記各発話様式間におけるデータ量の違いを補正しながら、上記学習データに対応する音声パターンモデルのパラメータを求める重み付き学習手段とを備えた音声パターンモデル学習装置。

【請求項2】 音声パターンモデルは、隠れマルコフモデルであり、

重み付き学習手段は、重み計算手段が算出した重み係数を学習データから算出した遷移回数期待値に乗じて、上記学習データの各発話様式間におけるデータ量の違いを補正した重み付き遷移回数期待値とし、この重み付き遷移回数期待値を用いて上記隠れマルコフモデルのパラメータを求めることを特徴とする請求項1記載の音声パターンモデル学習装置。

【請求項3】 異なる発話様式で入力された音声データの音響的特徴を表す複数種類の学習データを格納する学習データ記憶手段と、

上記学習データの発話様式ごとに応する音声パターンモデルのパラメータを求める発話様式別音声パターンモデル学習手段と、

この発話様式別音声パターンモデル学習手段が求めた発話様式ごとのパラメータに対応する発話様式別音声パターンモデルを用いて、上記学習データ記憶手段が格納する上記学習データをクラスタリングし、上記各学習データが属する発話様式のクラスタを決定するクラスタリング手段と、

このクラスタリング手段がクラスタリングした各発話様式のクラスタに属する学習データのデータ量の逆数に、これらの総和が1となるように正規化したものを、上記学習データに対する上記発話様式のクラスタごとのデータ量に応じたクラスタ重み係数として算出するクラスタ重み計算手段と、

このクラスタ重み計算手段が算出した上記クラスタ重み係数を用いて上記学習データの上記各発話様式のクラスタ間におけるデータ量の違いを補正しながら、上記学習データに対応する音声パターンモデルのパラメータを求めるクラスタ重み付き学習手段とを備えた音声パターンモデル学習装置。

【請求項4】 音声パターンモデルは、隠れマルコフモデルであり、

クラスタ重み付き学習手段は、クラスタ重み計算手段が

算出したクラスタ重み係数を学習データから算出した遷移回数期待値に乗じて、上記学習データの各発話様式のクラスタ間におけるデータ量の違いを補正したクラスタ重み付き遷移回数期待値とし、このクラスタ重み付き遷移回数期待値を用いて上記隠れマルコフモデルのパラメータを求める特徴とする請求項3記載の音声パターンモデル学習装置。

【請求項5】 異なる発話様式で入力された音声データの音響的特徴を表す複数種類の学習データにおける上記発話様式ごとの各データ量の逆数に、これらの総和が1となるように正規化したものを、上記学習データに対する上記発話様式ごとのデータ量に応じた重み係数として算出する重み計算ステップと、この重み計算ステップで算出した重み係数を用いて上記学習データの上記各発話様式間におけるデータ量の違いを補正しながら、上記学習データに対応する音声パターンモデルのパラメータを求める重み付き学習ステップとを備えた音声パターンモデル学習方法。

【請求項6】 音声パターンモデルは、隠れマルコフモデルであり、

重み付き学習ステップにて、重み計算ステップで算出した重み係数を学習データから算出した遷移回数期待値に乗じて、上記学習データの各発話様式間におけるデータ量の違いを補正した重み付き遷移回数期待値とし、この重み付き遷移回数期待値を用いて上記隠れマルコフモデルのパラメータを求める特徴とする請求項5記載の音声パターンモデル学習方法。

【請求項7】 異なる発話様式で入力された音声データの音響的特徴を表す複数種類の学習データの上記発話様式ごとに応する音声パターンモデルのパラメータを求める発話様式別音声パターンモデル学習ステップと、この発話様式別音声パターンモデル学習ステップにて求めた発話様式ごとのパラメータに対応する発話様式別音声パターンモデルを用いて上記学習データをクラスタリングし、上記各学習データが属する発話様式のクラスタを決定するクラスタリング手段と、

このクラスタリングステップでクラスタリングした各発話様式のクラスタに属する学習データのデータ量の逆数に、これらの総和が1となるように正規化したものを、上記学習データに対する上記発話様式のクラスタごとのデータ量に応じたクラスタ重み係数として算出するクラスタ重み計算ステップと、

このクラスタ重み計算ステップで算出した上記クラスタ重み係数を用いて上記学習データの上記各発話様式のクラスタ間におけるデータ量の違いを補正しながら、上記学習データに対応する音声パターンモデルのパラメータを求めるクラスタ重み付き学習ステップとを備えた音声パターンモデル学習方法。

【請求項8】 音声パターンモデルは、隠れマルコフモデルであり、

クラスタ重み付き学習ステップにて、クラスタ重み計算ステップで算出したクラスタ重み係数を学習データから算出した遷移回数期待値に乗じて、上記学習データの各発話様式のクラスタ間におけるデータ量の違いを補正したクラスタ重み付き遷移回数期待値とし、このクラスタ重み付き遷移回数期待値を用いて上記隠れマルコフモデルのパラメータを求める特徴とする請求項7記載の音声パターンモデル学習方法。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】この発明は朗読調の丁寧な発声の音声や自由発話音声のように発話速度がはやく曖昧な音声などのように発話様式の異なる複数種類の発話の音響特徴を適切にモデル化する音声パターンモデル学習装置及び音声パターンモデル学習方法に関するものである。

【0002】

【従来の技術】音声認識は、一般に音声データを音響分析して得られる音声データの音響的特徴量である学習データに相当する特徴ベクトルの時系列と、その特徴ベクトルの時系列のパターンをモデル化した音声パターンモデルとのパターンマッチングを行うことにより実現される。この音声パターンモデルとしては、隠れマルコフモデル(Hidden Markov Model 以下、HMMと称する)が用いられることが多い。このHMMでは、大量の学習データを用意することができれば、音声データにおける特徴ベクトルの時系列のパターンを精度よくモデル化することができる。

【0003】しかしながら、認識対象とする音声データが、学習データとは異なる発話様式で入力されたものであると、認識精度が低下するという問題点があった。特に、連続音声認識においては、テキストを読み上げた音声でHMMを学習しても、対話音声のように自由に発声した音声に対してはロバストなモデルにはならず、充分な認識率が得られないことが知られている。この問題に対処するために、Richard P. Lippmann, et al., "Multi-Style Training for Robust Isolated-Word Speech Recognition", IEEE International Conference on Acoustics, Speech, and Signal Processing, vol. 2 of 4, pp. 705-708, 1987(以下、文献1と称する)では、様々な発話様式で入力された音声データを用いてHMMを学習する方法を提案している。なお、上記文献1では単語音声のモデル化に適用した場合を説明しているが、全く同じ技術を連続音声のモデル化に適用することができる。

【0004】ここで、連続音声のモデル化に文献1の技術を使用し、テキストを読み上げたような丁寧な音声

と、対話音声のように自由に発声した音声の両方に対してロバストなHMMを学習する方法を説明する。本例では学習するHMMは連続分布型のHMMであり、1個のHMMで1個の音素をモデル化するものとする。また、HMMのトロポジは、例えば図6に示すように4状態のleft-to-right型とする。HMMは、遷移確率 a_{ij} と出力確率 $b_{ij}(X)$ とをパラメータとして持つ。ここで、添字 i, j は、HMMにおける音声データの音響的特徴の状態が状態 i から j に遷移することを示しており、本例では4状態のHMMなので、 $i = (1, 2, 3), j = (i, i+1)$ となる。また、遷移確率 a_{ij} は状態 i から状態 j への遷移が起きる確率であり、出力確率 $b_{ij}(X)$ は状態 i から j への遷移の際に特徴ベクトル X が出力される確率密度であり、多次元正規分布で表現される。即ち、 $b_{ij}(X)$ を表現するパラメータが平均値 m_{ij} と分散 v_{ij} であることから、HMMのパラメータは、遷移確率 a_{ij} と出力確率を表現する平均値 m_{ij} 及び分散 v_{ij} となる。このようなHMMのパラメータを求めることがHMMの学習といい、音声データを音響分析して得られる特徴ベクトルの時系列などのHMMの学習に用いるデータを学習データと定義する。

【0005】図7は文献1に開示された技術を用いた従来の音声パターンモデル学習装置の構成を示すブロック図である。図において、100は読み上げ音声データ100aや自由発話音声データ100bなどのHMMの学習データ110を格納する学習データメモリ、100aはテキストを読み上げたような丁寧な音声で入力された音声データを音響分析して得られる読み上げ音声データで、100bは対話音声のように自由に発声した音声で入力された音声データを音響分析して得られる自由発話音声データである。110は学習データで、読み上げ音声データ100aや自由発話音声データ100bのうち、HMMの学習時に学習手段120によって学習データメモリ100から適宜読み出されてHMMの学習に使用されるデータである。120は学習データメモリ100内に格納される音声データに対するHMMの学習を行う学習手段、130は学習手段120が学習したHMMのパラメータ、140は学習手段120が学習したHMMのパラメータ130を格納するモデルメモリである。また、学習データメモリ100には読み上げ音声データ100aと自由発話音声データ100b、及び、学習対象とする音素名の一覧を記した音素テーブルが格納されている。図8は上述した音素テーブルの一例を示す図である。図に示すように、認識対象の音声データを音素ごとに分割し、これらに通し番号を付して管理している。

【0006】先ず、学習データメモリ100内の各データの概要について説明する。読み上げ音声データ100aは、多様な音素のコンテキストを含んだ文章を多数の話者が読み上げて入力した音声データを音響分析して得

られる特徴ベクトルの時系列と発話内容を示す音素表記から構成される。具体的には、読み上げによって入力された音声データの音声波形を音響分析して得られる特徴ベクトルXの時系列を音素区間ごとに切り出したトークン $X^{(k)}_1, X^{(k)}_2, X^{(k)}_3, \dots, X^{(k)}_{T(k)}$ （肩の添字(k)はk番目のトークンであることを意味する）の集合と、各トークンのトークン番号と音素表記を記した読み上げ音声トークンテーブルである。図9は上述した読み上げ音声トークンテーブルを示す図である。図に示すように、各トークンのトークン番号は1番から順番に付すものとする。また、音響分析は、例えばLPC分析とし、特徴ベクトルXはLPCケプストラムである。さらに、音素区間ごとの切り出しへは、例えば人間がスペクトログラムを観察して行うものとする。

【0007】自由発話音声データ100bは、多様な話題に対する人対人の対話音声を音響分析して得られる特徴ベクトルXの時系列と発話内容を示す音素表記から構成される。具体的には、自由発話によって入力された音声データの音声波形を音響分析して得られる特徴ベクトルの時系列を音素区間ごとに切り出したトークン $X^{(k)}_1, X^{(k)}_2, X^{(k)}_3, \dots, X^{(k)}_{T(k)}$ （肩の添字(k)はk番目のトークンであることを意味する）の集合と、各トークンのトークン番号と音素表記を記した自由発話音声トークンテーブルである。図10は上述した自由発話音声データトークンテーブルを示す図である。図に示すように、自由発話音声データ100b中のトークン番号は、読み上げ音声トークンテーブルの最終番号に続く番号を付与する。音響分析は読み上げ音声データと同様に例えばLPC分析とし、特徴ベクトルXはLPCケプストラムである。音素区間ごとの切り出しへは、読み上げ音声データ100aと同様に、例えば人間がスペクトログラムを観察して行うものとする。読み上げ音声データ100aは、テキストを読み上げているので比較的丁寧で明瞭な発声であるのに

$$a_{ij} = \frac{\sum_{k=1}^K \sum_{t=1}^{T(k)} \gamma^{(k)}(i, j, t)}{\sum_{k=1}^K \sum_{j=i}^{i+1} \sum_{t=1}^{T(k)} \gamma^{(k)}(i, j, t)} \quad (1)$$

【数2】

$$m_{ij} = \frac{\sum_{k=1}^K \sum_{t=1}^{T(k)} \gamma^{(k)}(i, j, t) * X_t^{(k)}}{\sum_{k=1}^K \sum_{t=1}^{T(k)} \gamma^{(k)}(i, j, t)} \quad (2)$$

【数3】

対し、自由発話音声データ100bは人対人の自然な対話音声であるため音素の特徴ベクトルの変形が激しいのが特徴である。

【0008】次にHMMを学習する動作について説明する。学習手順1 学習手段120は、学習データメモリ100が保持する音素テーブルを読み込み、この音素テーブルの記述にしたがって、学習対象とする音素を選択する。学習手段120は、音素テーブルが、例えば図8のように記述されている場合、先頭の音素である/a/を学習対象として選択する。

【0009】学習手順2 学習手段120は、選択した音素と一致する音素表記を持つ全てのトークンの特徴ベクトルの時系列を学習データメモリ100から読み込む。このとき、読み上げ音声データ100aのトークンと自由発話音声データ100bのトークンを両方読み込む。ここで、読み込んだトークンの総数をKとして、読み込んだ各トークン $X^{(k)}_1, X^{(k)}_2, X^{(k)}_3, \dots, X^{(k)}_{T(k)}$ に対する遷移回数期待値 $\gamma^{(k)}(i, j, t)$ を計算する。但し、肩の添字(k)は読み込んだ全てのトークン中でk(k=1, 2, 3, ..., K)番目のトークンであることを意味する。Xはトークンを構成する特徴ベクトル、T(k)はk番目のトークンを構成する特徴ベクトルの数とする。また、(i, j, t)はトークンのt番目の特徴ベクトル $X^{(k)}_t$ でHMMの状態iから状態jへ遷移したことを示すものとする。この遷移回数期待値 $\gamma^{(k)}(i, j, t)$ は、例えばフォワード・バックワードアルゴリズムを用いて計算する。

【0010】次に、遷移回数期待値 $\gamma^{(k)}(i, j, t)$ を用いて、(1)～(3)式によってHMMのパラメータである遷移確率 a_{ij} 、平均値 m_{ij} 及び分散 v_{ij} を学習する。HMMは、図6に示したように、4状態なのでi=(1, 2, 3)、j=(i, i+1)である。

【数1】

$$v_{ij} = \frac{\sum_{k=1}^K \sum_{t=1}^{T^{(k)}} \gamma^{(k)}(i, j, t) * (m_{ij} - X_i^{(k)})^2}{\sum_{k=1}^K \sum_{t=1}^{T^{(k)}} \gamma^{(k)}(i, j, t)} \quad (3)$$

学習対象の音素に対するHMMの学習を終了すると、学習手段120は学習を終了したモデルのパラメータ130である、遷移確率 a_{ij} 、平均値 m_{ij} 、分散 v_{ij} 、及び、その音素表記をモデルメモリ140に送出する。モデルメモリ140は、学習を終了した上記モデルのパラメータ130及びその音素表記を保持する。

【0011】学習手順3 学習手段120は、学習データメモリ100が保持する音素テーブルに存在する全ての音素に対してモデルの学習が終了するまで、上述した手順で学習対象とする音素を音素テーブルに記述されている順序にしたがって選択し、上記学習手順2を繰り返す。以上で音素モデルの学習を終了する。

【0012】このように文献1に示された従来技術では、音響的特徴の異なる種々の発話様式の学習データを用いることにより、種々の発話様式の音声に対してロバストなHMMを得ることを目的としている。本例では読み上げ音声データ100aと自由発話音声データ100bとの両方に対してロバストなHMMを得ることを目的としている。

【0013】

【発明が解決しようとする課題】従来の音声パターンモデル学習装置は以上のように構成されているので、各音素の全てに対して読み上げ音声データ100aと自由発話音声データ100bとの学習データ量が同量程度でなければ、両方の発話様式に対して等しくロバストなHMMを得られないという課題があった。

【0014】上記課題について具体的に説明する。自由発話音声データ100bは、一般的に多様な話題に対する人対人の対話音声を音響分析して得ることからデータ収集が困難であり、読み上げ音声データ100aと比較してHMMの学習に使用する学習データ110のデータ量が少なくなる。また、読み上げ音声データ100aは、発話内容を完全に指定して音声を収録することができるが、自由発話音声データ100bは、一般的に人対人の自由発話音声データ等を収録するので発話内容を完全に指定することができない。従って、読み上げ音声とは音素の出現頻度が異なったものになる。HMMの学習は、(1)～(3)式に示すように最尤推定に基づいていたことから、データ量と音素の出現頻度が異なる学習データを両方用いて、文献1に示された従来技術によってHMMを学習すると、読み上げ音声の特徴と自由発話音声データとの音響的特徴が均等にモデル化されないという不具合があった。即ち、読み上げ音声データ100aの方が学習データ量が多くなり、HMMの学習結果が読み上げ音声データ100aの音響的特徴に近いモデル

となり、自由発話音声データ100bの音響的特徴はモデル化されにくくなる。

【0015】一方、上記課題を解決する従来の技術として、特開平4-326400号公報（以下、文献2と称する）に開示される音響モデル構成方法がある。この文献2に開示される技術は、読み上げ音声データと自由発話音声データとで、別々にHMMを学習する（文献1による技術では、上述したように学習手段120が読み上げ音声データ100aと自由発話音声データ100bとを両方読み込んで、両者を考慮してHMMの学習を行っている）。このあと、音声認識時に各HMMの尤度を計算して加重平均を取る。この加重平均により算出された尤度を最適な尤度として設定し、この尤度に対応するパラメータによるHMMを、音声データに最適化された音声パターンモデルとして決定する。しかしながら、上記文献2による技術を使用した音声パターン学習装置では、パラメータを求めるHMMの数が読み上げ音声データと自由発話音声データとで2倍になり、さらに、HMMパラメータの記憶領域とHMMの尤度計算の演算量も2倍になるという課題があった。これにより、上記演算量と記憶領域とを提供できる程のハードウェア資源を要することからコスト的に不利であった。

【0016】この発明は上記のような課題を解決するためになされたもので、読み上げ音声と自由発話音声との学習データ量が異なる場合でも両方の発話様式に対してHMMの数を増加させることなく、ロバストなHMMを求めることができる音声パターンモデル学習装置を得ることを目的とする。

【0017】

【課題を解決するための手段】この発明に係る音声パターンモデル学習装置は、異なる発話様式で入力された音声データの音響的特徴を表す複数種類の学習データを格納する学習データ記憶手段と、複数種類の学習データにおける発話様式ごとの各データ量の逆数に、これらの総和が1となるように正規化したものを、学習データに対する発話様式ごとのデータ量に応じた重み係数として算出する重み計算手段と、この重み計算手段が算出した重み係数を用いて学習データの各発話様式間におけるデータ量の違いを補正しながら、学習データに対応する音声パターンモデルのパラメータを求める重み付き学習手段とを備えるものである。

【0018】この発明に係る音声パターンモデル学習装置は、音声パターンモデルが隠れマルコフモデルであり、重み付き学習手段は、重み計算手段が算出した重み係数を学習データから算出した遷移回数期待値に乘じ

て、学習データの各発話様式間におけるデータ量の違いを補正した重み付き遷移回数期待値とし、この重み付き遷移回数期待値を用いて隠れマルコフモデルのパラメータを求めるものである。

【0019】この発明に係る音声パターンモデル学習装置は、異なる発話様式で入力された音声データの音響的特徴を表す複数種類の学習データを格納する学習データ記憶手段と、学習データの発話様式ごとに応する音声パターンモデルのパラメータを求める発話様式別音声パターンモデル学習手段と、この発話様式別音声パターンモデル学習手段が求めた発話様式ごとの音声パターンモデルのパラメータに対応する発話様式別音声パターンモデルを用いて、学習データ記憶手段が格納する学習データをクラスタリングし、各学習データが属する発話様式のクラスタを決定するクラスタリング手段と、このクラスタリング手段がクラスタリングした各発話様式のクラスタに属する学習データのデータ量の逆数に、これらの総和が1となるように正規化したものを、学習データに対する発話様式のクラスタごとのデータ量に応じたクラスタ重み係数として算出するクラスタ重み計算手段と、このクラスタ重み計算手段が算出したクラスタ重み係数を用いて学習データの各発話様式のクラスタ間におけるデータ量の違いを補正しながら、学習データに対する音声パターンモデルのパラメータを求めるクラスタ重み付き学習手段とを備えるものである。

【0020】この発明に係る音声パターンモデル学習装置は、音声パターンモデルが隠れマルコフモデルであり、クラスタ重み付き学習手段は、クラスタ重み計算手段が算出したクラスタ重み係数を学習データから算出した遷移回数期待値に乗じて、学習データの各発話様式のクラスタ間におけるデータ量の違いを補正したクラスタ重み付き遷移回数期待値とし、このクラスタ重み付き遷移回数期待値を用いて隠れマルコフモデルのパラメータを求めるものである。

【0021】この発明に係る音声パターンモデル学習方法は、異なる発話様式で入力された音声データの音響的特徴を表す複数種類の学習データにおける発話様式ごとの各データ量の逆数にこれらの総和が1となるように正規化したものを、学習データに対する発話様式ごとのデータ量に応じた重み係数として算出する重み計算ステップと、この重み計算ステップで算出した重み係数を用いて学習データの各発話様式間におけるデータ量の違いを補正しながら、学習データに対する音声パターンモデルのパラメータを求める重み付き学習ステップとを備えるものである。

【0022】この発明に係る音声パターンモデル学習方法は、音声パターンモデルが隠れマルコフモデルであり、重み付き学習ステップにて、重み計算ステップで算出した重み係数を学習データから算出した遷移回数期待値に乗じて、学習データの各発話様式間におけるデータ

量の違いを補正した重み付き遷移回数期待値とし、この重み付き遷移回数期待値を用いて隠れマルコフモデルのパラメータを求めるものである。

【0023】この発明に係る音声パターンモデル学習方法は、異なる発話様式で入力された音声データの音響的特徴を表す複数種類の学習データの発話様式ごとに応する音声パターンモデルのパラメータを求める発話様式別音声パターンモデル学習ステップと、この発話様式別音声パターンモデル学習ステップにて求めた発話様式ごとの音声パターンモデルのパラメータに対応する発話様式別音声パターンモデルを用いて学習データをクラスタリングして、各学習データが属する発話様式のクラスタを決定するクラスタリングステップと、このクラスタリングステップでクラスタリングした各発話様式のクラスタに属する学習データのデータ量の逆数に、これらの総和が1となるように正規化したものを、学習データに対する発話様式のクラスタごとのデータ量に応じたクラスタ重み係数として算出するクラスタ重み計算ステップと、このクラスタ重み計算ステップで算出したクラスタ重み係数を用いて学習データの各発話様式のクラスタ間におけるデータ量の違いを補正しながら、学習データに対する音声パターンモデルのパラメータを求めるクラスタ重み付き学習ステップとを備えるものである。

【0024】この発明に係る音声パターンモデル学習方法は、音声パターンモデルが隠れマルコフモデルであり、クラスタ重み付き学習ステップにて、クラスタ重み計算ステップで算出したクラスタ重み係数を学習データから算出した遷移回数期待値に乗じて、学習データの各発話様式のクラスタ間におけるデータ量の違いを補正したクラスタ重み付き遷移回数期待値とし、このクラスタ重み付き遷移回数期待値を用いて隠れマルコフモデルのパラメータを求めるものである。

【0025】

【発明の実施の形態】以下、この発明の実施の一形態を説明する。

実施の形態1. 図1はこの発明の実施の形態1による音声パターンモデル学習装置の構成を示すブロック図である。図において、1は読み上げ音声データ1-1や自由発話音声データ1-2などのHMMの学習データ2を格納する学習データメモリ(学習データ記憶手段)、1-1はテキストを読み上げたような丁寧な音声で入力された音声データを音響分析して得られる読み上げ音声データ(学習データ)で、1-2は対話音声のように自由に発声した音声で入力された音声データを音響分析して得られる自由発話音声データ(学習データ)である。具体的に読み上げ音声データ1-1及び自由発話音声データ1-2を説明すると、読み上げ音声データ1-1は、読み上げ音声によって入力された音声データの音声波形を音響分析して得られる特徴ベクトルXの時系列を音素区間にごとに切り出したトークンの集合と、各トークンのト

ークン番号と音素表記を記した読み上げ音声トークンテーブルを指している。また、自由発話音声データ1-2は、自由発話によって入力された音声データの音声波形を音響分析して得られる特徴ベクトルの時系列を音素区間にごとに切り出したトークンの集合と、各トークンのトークン番号と音素表記を記した自由発話音声トークンテーブルを指している。2は学習データで、読み上げ音声データ1-1や自由発話音声データ1-2のうち重み計算手段6や重み付き学習手段9によって学習データメモリ1から適宜読み出されてHMMの学習に使用されるデータである。

【0026】6は読み上げ音声データ1-1や自由発話音声データ1-2から読み出した学習データ2の各データ量から学習データ2に対する発話様式（読み上げ音声、自由発話音声）ごとのデータ量に応じた重み係数を算出する重み計算手段で、7は重み計算手段6が算出した重み係数である。8は重み計算手段6が算出した重み係数7を格納する重み計算結果メモリ、9は重み計算手段6が算出した重み係数7を用いて、学習データ2の各発話様式（読み上げ音声、自由発話音声）間におけるデータ量の違いを補正しながら、学習データ2に対する音声パターンモデルのパラメータを求める重み付き学習手段、10は重み付き学習手段9によって学習したHMMのパラメータ（学習データ2に対応する音声パターンモデルのパラメータ）を含むHMMパラメータ情報、11は重み付き学習手段9が学習データ2の各発話様式（読み上げ音声、自由発話音声）間におけるデータ量の違いを補正しながら求めたHMMパラメータ情報10を格納する重み付きモデルメモリである。なお、この実施の形態1で学習するHMMは、上述した従来技術と同様に連続分布型のHMMとし、1個のHMMで1個の音素をモデル化するものとする。HMMのトポロジーは図6に示すような4状態のleft-right型とする。

【0027】次に動作について説明する。この実施の形

$$WR = \frac{CS}{CR + CS}$$

【数5】

$$WS = \frac{CR}{CR + CS}$$

重み計算手段6は、当該音素の音素表記である/a/とともに、算出された重み係数WR、WSを重み計算結果メモリ8に送出する。これにより、重み計算結果メモリ8は音素表記である/a/と重み係数WR、WSとを保持する。

【0030】重み係数の計算手順3 重み計算手段6は、図8に示した音素テーブルの表記の順番、即ち、/i/～/z/の順番に上述した重み係数の計算手順2を

態1による音声パターンモデル学習装置の学習手順は、（1）重み係数の計算（重み計算ステップ）、（2）重み付きモデルの学習（重み付き学習ステップ）の2段階に分かれる。先ず、重み係数の計算手順について説明する。重み係数の計算を行う場合、学習データメモリ1の出力端子Aを重み計算手段6の入力端子A1に接続する。この接続状態で重み係数の計算を以下の手順で行う。

【0028】重み係数の計算手順1 重み計算手段6は、上記接続経路を介して学習データメモリ1が保持する音素テーブルを読み込み、この音素テーブルに記述の音素に付された順番に従って、重み係数の計算対象とする音素を選択する。ここで、音素テーブルが、例えば図8のように記述されている場合、通し番号が先頭の音素である/a/を重み係数の計算対象として選択する。

【0029】重み係数の計算手順2 次に、重み計算手段6は、学習データメモリ1が保持する読み上げ音声トークンテーブル（学習データ2）と自由発話音声トークンテーブル（学習データ2）とに記載されているトークンのうち、重み係数の計算対象音素として選択した音素表記を持つトークンの数を上記両トークンテーブルで別々に数え上げる。ここで、読み上げ音声トークンテーブルに対するトークン（学習データ2）の数え上げ数をCR、自由発話音声トークンテーブルに対するトークン（学習データ2）の数え上げ数をCSとすると、読み上げ音声データ1-1に対する重み係数WRと自由発話音声データデータ1-2に対する重み係数WSとは、下記（4）、（5）式にしたがって計算される。これら（4）、（5）式から分かるとおり、読み上げ音声データ1-1に対する重み係数WRと自由発話音声データデータ1-2に対する重み係数WSとの比は、読み上げ音声データ1-1に属するトークンの数CRと自由発話音声データデータ1-2に属するトークンの数CSの比に反比例する値となっている。

【数4】

(4)

(5)

繰り返し、各音素ごとに読み上げ音声データ1-1に対する重み係数WRと自由発話音声データデータ1-2に対する重み係数WSを計算して、当該音素の音素表記とともに重み係数WR、WSを重み計算結果メモリ8に送出する。重み計算結果メモリ8は音素表記と重み係数WR、WSとを保持する。図2は実施の形態1による音声パターンモデル学習装置における重み計算結果メモリに保持された重み係数を示す図である。図において、音素

/a/に対する重み係数値は、WR=0.25、WS=0.75となっているが、これは音素/a/に対する読み上げ音声データ1-1のトークン数と自由発話音声データ1-2のトークン数の比が0.75:0.25、即ち、読み上げ音声データ1-1のトークン数が自由発話音声データ1-2のトークン数の3倍であることを意味する。このようにして学習データに対する発話様式ごとのデータ量に応じた重み係数の算出が完了する。

【0031】次に重み付きモデルの学習手順を説明する。重み付きモデルの学習を開始する前に、学習データメモリ1の出力端子Aを重み付き学習手段9の入力端子A3に接続する。この接続状態で重み付きモデルの学習を以下の手順で行う。

【0032】重み付きモデルの学習手順1 重み付き学習手段9は、上記接続経路を介して学習データメモリ1が保持する音素テーブルを読み込み、この音素テーブルの記述の音素に付された順番に従って、学習対象とする音素を選択する。ここで、音素テーブルが、例えば図8のように記述されている場合、通し番号が先頭の音素である/a/を学習対象として選択する。

【0033】重み付きモデルの学習手順2 次に、重み付き学習手段9は、上述のようにして選択した音素と一致する音素表記を持つ全てのトークンの特徴ベクトルの時系列である学習データ2を学習データメモリ1から読み込む。この際、読み上げ音声データ1-1中のトークン

$$\gamma_w^{(k)}(i, j, t)$$

$$= \begin{cases} WR * \gamma^{(k)}(i, j, t) & (k\text{番目のトークンが読み上げ音声データ中のトークンの場合}) \\ WS * \gamma^{(k)}(i, j, t) & (k\text{番目のトークンが自由発話音声データ中のトークンの場合}) \end{cases}$$

重み付き学習手段9は、算出した重み付き遷移回数期待値 $\gamma^{(k)}_w(i, j, t)$ を用いて下記(7)～(9)式に従って、HMMのパラメータを学習する。こ

$$a_{ij}^w = \frac{\sum_{k=1}^K \sum_{t=1}^{T(k)} \gamma_w^{(k)}(i, j, t)}{\sum_{k=1}^K \sum_{j=i}^{i+1} \sum_{t=1}^{T(k)} \gamma_w^{(k)}(i, j, t)} \quad (7)$$

【数8】

$$m_{ij}^w = \frac{\sum_{k=1}^K \sum_{t=1}^{T(k)} \gamma_w^{(k)}(i, j, t) * X_t^{(k)}}{\sum_{k=1}^K \sum_{t=1}^{T(k)} \gamma_w^{(k)}(i, j, t)} \quad (8)$$

【数9】

ンと自由発話音声データ1-2中のトークンを両方読み込む。読み込んだトークンの総数をKとする。また、重み付き学習手段9は、重み計算結果メモリ8から上記学習手順1で選択した音素に対する読み上げ音声データのWRと自由発話音声データデータの重み係数WSとを読み込む。

【0034】重み付きモデルの学習手順3 重み付き学習手段9は、読み込んだ各トークン $X^{(k)}_1, X^{(k)}_2, X^{(k)}_3, \dots, X^{(k)}_T, (k)$ に対する遷移回数期待値 $\gamma^{(k)}(i, j, t)$ を計算する。ここで、肩の添字(k)は読み込んだ全てのトークン中でk(k=1, 2, 3, ..., K)番目のトークンであることを意味する、Xはトークンを構成する特徴ベクトル、T(k)はk番目のトークンを構成する特徴ベクトルの数とする。また、(i, j, t)は、トークンのt番目の特徴ベクトル $X^{(k)}_t$ を出力して、HMMの状態iから状態jへ遷移したことを示すものとする。この遷移回数期待値 $\gamma^{(k)}(i, j, t)$ は、従来技術と同様に例えばフォワード・バックワードアルゴリズムを用いて計算することができる。重み付き学習手段9は、上記遷移回数期待値 $\gamma^{(k)}(i, j, t)$ を用いて重み付き遷移回数期待値 $\gamma^{(k)}_w(i, j, t)$ を(6)式に従って算出する。

【数6】

(6)

ここで、HMMは、図6に示すような4状態とすることから、i=(1, 2, 3)、j=(i, i+1)となる。

【数7】

(7)

(8)

$$v_{ij}^w = \frac{\sum_{k=1}^K \sum_{t=1}^{T^{(k)}} \gamma_w^{(k)}(i, j, t) * (m_{ij}^w - X_i^{(k)})^2}{\sum_{k=1}^K \sum_{t=1}^{T^{(k)}} \gamma_w^{(k)}(i, j, t)} \quad (9)$$

【0035】学習を終了すると、重み付き学習手段9は、遷移確率 a_{ij}^w 、平均値 m_{ij}^w 、分散 v_{ij}^w 及びその音素表記を重み付きモデルメモリ11に送出する。重み付きモデルメモリ11は、学習を終了した上記HMMのパラメータとその音素表記とをHMMパラメータ情報10として保持する。

【0036】重み付きモデルの学習手順4 重み付き学習手段9は、学習データメモリ1が保持する音素テーブルに存在する全ての音素に対してHMMの学習が終了するまで学習対象とする音素を音素テーブルに記述されている順序にしたがって選択し、上記重み付きモデルの学習手順2～3を繰り返す。

【0037】ここで、実施の形態1と従来技術との違いについて説明する。(7)～(9)式において、通常の遷移回数期待値 $\gamma^{(k)}(i, j, t)$ の代わりに重み付き遷移回数期待値 $\gamma^{(k)}_w(i, j, t)$ を用いることである。この重み付き遷移回数期待値 $\gamma^{(k)}_w(i, j, t)$ は(6)式に示したとおり当該トークンが読み上げ音声データ1-1中のトークンの場合は重み係数WRをかけ、自由発話音声データ1-2中のトークンの場合には重み係数WSをかけて計算する。重み係数WRとWSとは、(4)、(5)式に示したとおり読み上げ音声データ1-1と自由発話音声データ1-2のトークン数に反比例する値となっている。従つて、重み付き遷移回数期待値 $\gamma^{(k)}_w(i, j, t)$ を用いることによって、読み上げ音声データ1-1に属するトークン数と自由発話音声データ1-2に属するトークン数との不均衡を補正し、両発話様式の音響的特徴を均等に反映したHMMを学習することができる。即ち、読み上げ音声と自由発話音声との両方に対してロバストなHMMを得ることができる。

【0038】以上のように、この実施の形態1によれば、異なる発話様式（読み上げ音声、自由発話音声）で入力された音声データの音響的特徴を表す読み上げ音声データ1-1、自由発話音声データ1-2から抽出した学習データ2の各データ量の逆数 $1/CR$ 、 $1/CS$ に、これらの総和が1となるように正規化したもの（各逆数 $1/CR$ 、 $1/CS$ に $CR \cdot CS / (CR + CS)$ を乗算する、(4)、(5)式参照）を、学習データ2に対する発話様式ごとのデータ量に応じた重み係数WR、WSとして算出し、この重み係数WR、WSを用いて学習データ2の各発話様式間におけるデータ量の違いを補正しながら、学習データ2に対応する音声パターンモデルのパラメータを求めるので、読み上げ音声データ1-1に属するトークン数と自由発話音声データ1-2

に属するトークン数との不均衡を補正し、両発話様式の音響的特徴を均等に反映した音声パターンモデルを学習することができ、読み上げ音声と自由発話音声との両方に対してロバストな音声パターンモデルを得ることができるという効果が得られる。

【0039】また、文献2による従来技術と異なり、読み上げ音声と自由発話音声とで別々に音声パターンモデルの学習をすることがないので、音声パターンモデルの数を増加させない。これにより、文献2による従来技術と比較して高性能なハードウェア資源を要する事なく、コスト的にも有利である。

【0040】また、この実施の形態1によれば、音声パターンモデルが隠れマルコフモデルであり、重み係数WR、WSを学習データ2から算出した遷移回数期待値 $\gamma^{(k)}(i, j, t)$ に乘じて、学習データ2の各発話様式間におけるデータ量の違いを補正した重み付き遷移回数期待値 $\gamma^{(k)}_w(i, j, t)$ とし、この重み付き遷移回数期待値 $\gamma^{(k)}_w(i, j, t)$ を用いて、隠れマルコフモデルのパラメータ遷移確率 a_{ij}^w 、平均値 m_{ij}^w 、分散 v_{ij}^w を求めるので、読み上げ音声データ1-1に属するトークン数と自由発話音声データ1-2に属するトークン数との不均衡を補正し、両発話様式の音響的特徴を均等に反映したHMMを学習することができ、読み上げ音声と自由発話音声との両方に対してロバストなHMMを得ることができるという効果が得られる。

【0041】なお、上記実施の形態1では、HMMで音素をモデル化する場合を説明したが、音節など他の音響単位でも構わない。

【0042】実施の形態2、この実施の形態2は、異なる発話様式で入力された音声データの音響的特徴を表す複数種類の学習データの発話様式ごとに応する音声パターンモデルのパラメータを求めて、これら発話様式ごとの音声パターンモデルのパラメータに対応する発話様式別音声パターンモデルを用いて学習データをクラスタリングして、各学習データが属する発話様式のグラスターを決定し、クラスタリングした各発話様式のグラスターに属する学習データのデータ量の逆数に、これらの総和が1となるように正規化したものを、学習データに対する発話様式のグラスターごとのデータ量に応じたグラスター重み係数として算出し、このグラスター重み係数を用いて学習データの各発話様式のグラスター間に応するデータ量の違いを補正しながら、学習データに対応する音声パターンモデルのパラメータを求めるものである。

【0043】図3はこの発明の実施の形態2による音声

パターンモデル学習装置の構成を示すブロック図である。図において、3は読み上げ音声データ1-1及び自由発話音声データ1-2からの学習データ2の発話様式ごとに応するHMMのパラメータを求める学習手段（発話様式別音声パターンモデル学習手段）である。12は学習手段3が読み上げ音声データ1-1を用いて学習したHMM（発話様式別音声パターンモデル）のパラメータを含むHMMパラメータ情報で、13は学習手段3からのHMMパラメータ情報12に対応するHMMを格納する読み上げ音声モデルメモリである。14は学習手段3が自由発話音声データ1-2を用いて学習したHMM（発話様式別音声パターンモデル）のパラメータを含むHMMパラメータ情報で、15は学習手段3からのHMMパラメータ情報14に対応するHMMを格納する自由発話音声モデルメモリである。16はクラスタリング手段で、学習手段3が発話様式別（読み上げ音声データ1-1、自由発話音声データ1-2）に求めたHMMを用いて学習データメモリ1が格納する学習データ2をクラスタリングして、各学習データ2が属する発話様式のクラスタを決定する。17はクラスタリング手段16が決定した学習データ2が属する発話様式のクラスタを各学習データ2ごとに応付けてデータであるクラスタリング結果、18はクラスタリング結果17を格納するクラスタリング結果メモリである。

【0044】19はクラスタリング結果メモリ18が格納するクラスタリング結果17から各発話様式のクラスタに属する学習データ2のデータ量を計算し、学習データ2に対する発話様式のクラスタごとのデータ量に応じたクラスタ重み係数を算出するクラスタ重み計算手段、20はクラスタ重み計算手段19が算出した学習データ2のクラスタ重み係数を各学習データ2ごとに応付けてデータであるクラスタ重み計算結果、21はクラスタ重み計算手段19からのクラスタ重み計算結果20を格納するクラスタ重み計算結果メモリである。22はクラスタ重み付き学習手段であって、クラスタ重み計算結果メモリ21から読み出したクラスタ重み計算結果20とクラスタリング結果メモリ18から読み出したクラスタリング結果17とを用いて学習データ2の各発話様式のクラスタ間におけるデータ量の違いを補正しながら、学習データ2に対するHMMを学習する。

【0045】23はクラスタ重み付き学習手段22によって学習したHMMのパラメータ（学習データ2に対応する音声パターンモデルのパラメータ）を含むHMMパラメータ情報、24はクラスタ重み付き学習手段22からのHMMパラメータ情報23に対応するHMMを格納するクラスタ重み付きモデルメモリである。また、この実施の形態2では学習するHMMは、上記実施の形態1と同様に連続分布型のHMMとし、1個のHMMで1個の音素をモデル化するものとする。さらに、HMMのトポロジーは、図6に示すように4状態のleft-to-right型とする。なお、図1と同一構成要素には同一符号を付して重複する説明を省略する。

【0046】次に動作について説明する。この実施の形態2による音声パターンモデル学習装置の学習手順は、

- (1) 読み上げ音声モデルの学習（発話様式別音声パターンモデル学習ステップ）、(2) 自由発話音声モデルの学習（発話様式別音声パターンモデル学習ステップ）、(3) 学習用トークンのクラスタリング（クラスタリングステップ）、(4) クラスタ重み係数の計算（クラスタ重み計算ステップ）、(5) クラスタ重み付きモデルの学習（クラスタ重み付き学習ステップ）の5段階に分かれる。

【0047】先ず、読み上げ音声モデルの学習手順を説明する。読み上げ音声モデルの学習を行う際は、学習データメモリ1の出力端子Aを学習手段3の入力端子A1に接続する。また、学習手段3の出力端子Bを読み上げ音声モデルメモリ13の入力端子B1に接続する。この接続状態で読み上げ音声モデルの学習を行う。

【0048】読み上げ音声モデル学習手順1 学習手段3は、上記接続経路を介して学習データメモリ1が保持する音素テーブルを読み込み、音素テーブルに記述された順序に従って学習対象とする音素を選択する。音素テーブルは、例えば図8のように記述されている場合、先頭の音素である/a/を学習対象として選択する。

【0049】読み上げ音声モデル学習手順2 学習手段3は、上述のようにして選択した音素と一致する音素表記を持つトークンの特徴ベクトルの時系列である学習データ2を学習データメモリ1から読み込む。このとき、読み上げ音声データ1-1のトークンのみ読み込む。次に、読み込んだ各トークンを用いて従来技術と同様にして、フォワード・バックワードアルゴリズムに基づいて遷移回数期待値 $\alpha_{ijk}(i, j, t)$ を計算し、(1)～(3)式によってHMMのパラメータである遷移確率 a_{ij} 、平均値 m_{ij} 及び分散 v_{ij} を学習する。

【0050】学習を終了すると、学習手段3は学習を終了したモデルのパラメータである、遷移確率 a_{ij} 、平均値 m_{ij} 、分散 v_{ij} 及びその音素表記からなるHMMパラメータ情報12を、読み上げ音声モデルメモリ13に送出する。読み上げ音声モデルメモリ13では、学習を終了した上記HMMのパラメータとその音素表記に対応するHMMを保持する。

【0051】読み上げ音声モデル学習手順3 学習手段3は、学習データメモリ1が保持する音素テーブルに記載された全ての音素に対してモデルの学習が終了するまで学習対象とする音素を音素テーブルに記述されている順序に従って選択し、上記読み上げ音声モデル学習手順2を繰り返す。このようにして読み上げ音声モデルの学習が完了すると、音素テーブルに存在する全ての音素に対しての読み上げ音声の音響的特徴をモデル化したHM

Mが、読み上げ音声モデルメモリ13に格納されることになる。

【0052】次に、自由発話音声データモデルの学習手順を説明する。自由発話音声モデルの学習を行う際は、学習データメモリ1の出力端子Aを学習手段3の入力端子A1に接続する。また、学習手段3の出力端子Bを自由発話音声モデルメモリ15の入力端子B2に接続する。この接続状態で自由発話音声モデルの学習を行う。この自由発話音声データモデルの学習手順は、上述した読み上げ音声モデルの学習手順において、読み上げ音声データ1-1のトークンを学習データ2として用いる代わりに自由発話音声データ1-2のトークンを用いてHMMを学習し、学習結果を読み上げ音声モデルメモリ13の代わりに自由発話音声モデルメモリ15に送出して格納することによってなされる。この結果、音素テーブルに存在する全ての音素に対しての自由発話音声の音響的特徴をモデル化したHMMが自由発話音声モデルメモリ15に格納されることになる。

【0053】次に、学習用トークンのクラスタリング手順について説明する。クラスタリングは学習データメモリ1の各トークンが読み上げ音声と自由発話音声データとのいずれに近いかをクラス分けするために行うもので、学習用トークン（学習データ2）のクラスタリングを行う際は、学習データメモリ1の出力端子Aをクラスタリング手段16の入力端子A2に接続する。具体的なクラスタリング手順を以下に示す。

【0054】クラスタリング手順1 クラスタリング手段16が、上記接続経路を介して学習データメモリ1が保持する音素テーブルを読み込み、音素テーブルに記述された順序に従って、クラスタリング対象とする音素を選択する。音素テーブルが図8のように記述されている場合、先頭の音素である/a/をクラスタリング対象として選択する。

【0055】クラスタリング手順2 クラスタリング手段16は、上述のようにして選択された音素をモデル化するHMMを、読み上げ音声モデルメモリ13及び自由発話音声モデルメモリ15の各々から読み込む。ここで、読み上げ音声モデルメモリ13から読み込んだHMMをHMMR、自由発話音声モデルメモリ15から読み込んだHMMをHMMSと記すことにする。

【0056】クラスタリング手順3 次に、クラスタリング手段16は、上述のようにして選択した音素と一致する音素表記を持つ全てのトークンの特徴ベクトルの時系列とトークン番号とからなる学習データ2を学習データメモリ1から読み込む。このとき、読み上げ音声データ1-1のトークンと自由発話音声データ学習データ1-2のトークンを両方読み込む。

【0057】クラスタリング手順4 クラスタリング手段16は、クラスタリング手順3で読み込んだトークン番号の小さいトークンから順番にクラスタリング手順2

で読み込んだHMMであるHMMRとHMMSとの各々に対する尤度計算を行う。この尤度計算には、例えばトレリス又はビタビアルゴリズムを用いる。また、読み上げ音声モデルメモリ13から読み込んだHMMであるHMMRに対する尤度をLR、自由発話音声モデルメモリ15から読み込んだHMMであるHMMSに対する尤度をLSとしたとき、LR \geq LSであれば、当該トークンは読み上げ音声に属することを意味する記号Rを当該トークン番号とともにクラスタリング結果メモリ18に送出する。逆に、LR<LSであれば、当該トークンは自由発話音声に属することを意味する記号Sを当該トークン番号とともにクラスタリング結果メモリ18に送出する。クラスタリング手段16は、読み込んだ全てのトークンに対してHMMRとHMMSとの各々に対する尤度を計算してクラスタリングを行い、当該トークン番号とともにクラスタリング結果である記号R又はSを、クラスタリング結果メモリ18に格納する。

【0058】クラスタリング手順5 クラスタリング手段16は、学習データメモリ1が保持する音素テーブルを参照し、学習データメモリ1に存在する全ての音素に対して音素テーブルに記述されている順序に従って選択し、クラスタリング手順2~4を繰り返す。

【0059】以上の操作によってクラスタリングが完了する。以上のように、クラスタリング手順1~5を行うことによって学習データメモリ1中の全てのトークンのクラスタリング結果17がクラスタリング結果メモリ18に保持されることになる。図4は実施の形態2による音声パターンモデル学習装置のクラスタリング結果メモリの内容を例示的に示す図である。図に示したように、読み上げ音声データ1-1中のトークンであるトークン番号1の/a/はクラスタリング結果がRであり、読み上げ音声にクラスタリングされていることがわかる。また、読み上げ音声データ1-1中のトークンであるトークン番号120の/a/はクラスタリング結果がSであり、自由発話音声にクラスタリングされている。これは、このトークンを含む発話が、読み上げ音声であっても自由発話音声に近かったことを意味している。

【0060】上記のようにクラスタリングを行う理由は、読み上げ音声データ1-1中のトークンであっても自由発話音声に近い発話もあり、逆に自由発話音声データ1-2中のトークンであっても読み上げ音声に近い発話もある。このため、クラスタリングを行う前の読み上げ音声データ1-1中のトークンと自由発話音声データ1-2中のトークンという分類では、必ずしも読み上げ音声の音響的特徴を持つトークンと自由発話音声データの音響的特徴を持つトークンとによるクラスに正確に分類されていないからである。

【0061】そこで、この実施の形態2のようにHMMRとHMMSの尤度に基づいて各トークンをクラスタリングすることにより、より正確に読み上げ音声の音響的

特徴を持つトーケンと自由発話音声の音響的特徴を持つトーケンとにクラス分類することができる。クラスの分類がより正確になることの利点は後述するように両クラスタの音響的特徴をより均等に反映したHMMを学習することができる。なお、クラスタリングの結果がRのクラスタを読み上げ音声クラスタ、クラスタリングの結果がSのクラスタを自由発話音声クラスタと呼ぶことにする。

【0062】次にクラスタ重み係数の計算手順について説明する。

クラスタ重み係数の計算手順1 クラスタ重み計算手段19は、出力端子A及び入力端子A2による経路を介して学習データメモリ1が保持する音素テーブルを読み込み、この音素テーブルに記述された順番に従ってクラスタ重み係数の計算対象とする音素を選択する。音素テーブルは、例えば図8のように記述されている場合、先頭の音素である/a/を重み係数の計算対象音素として選択する。

$$WR_c = \frac{CS_c}{CR_c + CS_c}$$

【数10】

$$WS_c = \frac{CR_c}{CR_c + CS_c}$$

クラスタ重み計算手段19は、クラスタ重み係数の計算対象として選択している当該音素の音素表記とともに重み係数WR_C、WS_Cをクラスタ重み計算結果メモリ21に送出する。クラスタ重み計算結果メモリ21は、クラスタ重み係数WR_C及びWS_Cを音素表記に対応させたデータとして保持する。

【0064】クラスタ重み係数の計算手順3 クラスタ重み計算手段19は、図8中の音素表記の順番、すなわち/i/～/z/の順番に上記クラスタ重み係数の計算手順2を繰り返し、各音素ごとに読み上げ音声クラスタに対する重み係数WR_Cと自由発話音声クラスタに対する重み係数WS_Cとを計算して、当該音素の音素表記とともに重み係数WR_C及びWS_Cを、クラスタ重み計算結果メモリ21に送出する。クラスタ重み計算結果メモリ21では、上記重み係数WR_C、WS_Cを音素表記に対応させたデータとして保持する。

【0065】以上の操作によってクラスタ重み係数の計算が完了する。図5は実施の形態2による音声パターンモデル学習装置のクラスタ重み計算結果メモリに保持された内容を示す図である。図に示すように、音素/a/に対する重み係数値はWR_C=0.2、WS_C=0.8で、上記実施の形態1の図2における音素/a/に対する重み係数値WR=0.25、WS=0.75と比較するとWR_C<WRとなっており、係数値が異なっている。これは、自由発話音声データ1-2のトーケンの幾

【0063】クラスタ重み係数の計算手順2 クラスタ重み計算手段19は、クラスタ重み係数の計算対象として選択した音素のトーケン（学習データ2）のクラスタリング結果17をクラスタリング結果メモリ18から読み込み、読み上げ音声クラスタに属する（即ち、クラスタリングの結果として記号Rが付与された）トーケンの数CR_Cと自由発話音声クラスタに属する（即ち、クラスタリングの結果として記号Sが付与された）トーケンの数CS_Cとを数える。そしてCR_CとCS_Cとを用いて、読み上げ音声クラスタに対する重み係数WR_Cと自由発話音声クラスタに対する重み係数WS_Cとを、下記(10)、(11)式に従って計算する。(10)、(11)式からわかるとおり、読み上げ音声クラスタに対する重み係数WR_Cと自由発話音声クラスタに対する重み係数WS_Cの比は、読み上げ音声クラスタに属するトーケンの数CR_Cと自由発話音声クラスタに属するトーケンの数CS_Cとの比に反比例する値となっている。

【数10】

(10)

(11)

つかが読み上げクラスタにクラスタリングされ、読み上げクラスタに属するトーケン数が増加したこと示している。

【0066】次に、クラスタ重み付きモデルの学習手順を説明する。クラスタ重み付きモデルの学習を行う際は、学習データメモリ1の出力端子Aをクラスタ重み付き学習手段22の入力端子A3に接続する。この接続状態でクラスタ重み付きモデルの学習を以下の手順で行う。

【0067】クラスタ重み付きモデル学習手順1 クラスタ重み付き学習手段22は、学習データメモリ1が保持する音素テーブルを読み込み、音素テーブルに記述にしたがって、学習対象とする音素を選択する。音素テーブルは、図8のように記述されている場合、先頭の音素である/a/を学習対象として選択する。

【0068】クラスタ重み付きモデル学習手順2 クラスタ重み付き学習手段22は、上記接続経路を介して選択した音素と一致する音素表記を持つ全てのトーケンの特徴ベクトルの時系列とトーケン番号とを学習データ2として学習データメモリ1から読み込む。このとき、読み上げ音声データ1-1のトーケンと自由発話音声データ1-2のトーケンを両方読み込む。ここで、読み込んだトーケンの総数をKとする。また、クラスタ重み付き学習手段22は、クラスタ重み計算結果メモリ21からクラスタ重み付きモデル学習手順1で選択した音素に対

する読み上げ音声クラスタの重み係数 WR_C と自由発話音声クラスタの重み係数 WS_C とを読み込む。さらに、クラスタ重み付き学習手段22は、上述のようにして選択した音素と一致する音素表記を持つ全てのトークンのクラスタリング結果17をクラスタリング結果メモリ18から読み込む。

【0069】クラスタ重み付きモデル学習手順3 次に、クラスタ重み付き学習手段22は、読み込んだ各トークン $X^{(k)}_1, X^{(k)}_2, X^{(k)}_3, \dots, X^{(k)}_{T(k)}$ に対する遷移回数期待値 $\gamma^{(k)}(i, j, t)$ を計算する。ここで、肩の添字 (k) は読み込んだ全てのトークン中で k ($k = 1, 2, 3, \dots, K$) 番目のトークンであることを意味

$$\begin{aligned} \gamma_{WC}^{(k)}(i, j, t) \\ = \begin{cases} WR_C * \gamma^{(k)}(i, j, t) & (k\text{番目のトークンが読み上げ音声クラスタに属するとき}) \\ WS_C * \gamma^{(k)}(i, j, t) & (k\text{番目のトークンが自由発話音声クラスタに属するとき}) \end{cases} \end{aligned}$$

(12)

このクラスタ重み付き遷移回数期待値 $\gamma^{(k)}_{WC}(i, j, t)$ を用いて (13)～(15) 式によって HMM のパラメータを学習する。但し、HM

$$a_{ij}^{WC} = \frac{\sum_{k=1}^K \sum_{t=1}^{T(k)} \gamma_{WC}^{(k)}(i, j, k)}{\sum_{k=1}^K \sum_{j=t}^{T(k)} \gamma_{WC}^{(k)}(i, j, k)} \quad (13)$$

【数14】

$$m_{ij}^{WC} = \frac{\sum_{k=1}^K \sum_{t=1}^{T(k)} \gamma_{WC}^{(k)}(i, j, t) * X_i^{(k)}}{\sum_{k=1}^K \sum_{t=1}^{T(k)} \gamma_{WC}^{(k)}(i, j, t)} \quad (14)$$

【数15】

$$v_{ij}^{WC} = \frac{\sum_{k=1}^K \sum_{t=1}^{T(k)} \gamma_{WC}^{(k)}(i, j, k) * (m_{ij}^{WC} - X_i^{(k)})^2}{\sum_{k=1}^K \sum_{t=1}^{T(k)} \gamma_{WC}^{(k)}(i, j, k)} \quad (15)$$

学習を終了すると、クラスタ重み付き学習手段22は、学習を終了したモデルのパラメータである遷移確率 a_{ij}^{WC} 、平均値 m_{ij}^{WC} 、分散 v_{ij}^{WC} 、及びその音素表記からなる HMM パラメータ情報23をクラスタ重み付きモデルメモリ24に送出する。クラスタ重み付きモデルメモリ24は、学習を終了した HMM パラメータ情報23に対応する HMM を保持する。

【0070】クラスタ重み付きモデル学習手順4 クラスタ重み付き学習手段22は、学習データメモリ1が保持する音素テーブルに存在する全ての音素に対してモデルの学習が終了するまで学習対象とする音素を、音素テーブルに記述されている順序にしたがって選択し、上記クラスタ重み付き手順2～3を繰り返す。

【0071】ここで、この実施の形態2による音声バタ

する。Xはトークンを構成する特徴ベクトル、T(k)は k 番目のトークンを構成する特徴ベクトルの数とする。また (i, j, t) はトークンの t 番目の特徴ベクトル $X^{(k)}_t$ を出力して、HMM の状態 i から状態 j へ遷移したことを示すものとする。この遷移回数期待値 $\gamma^{(k)}(i, j, t)$ は、実施の形態1と同様に例えばフォワード・バックワードアルゴリズムを用いて計算することができる。このあと、遷移回数期待値 $\gamma^{(k)}(i, j, t)$ を用いて、クラスタ重み付き遷移回数期待値 $\gamma_{WC}^{(k)}(i, j, t)$ を下記(12)式にしたがって計算する。

【数12】

Mは、図6に示すような4状態であるので、 $i = (1, 2, 3)$ 、 $j = (i, i+1)$ となる。

【数13】

ーンモデル学習装置と上記実施の形態1による音声パターンモデル学習装置との違いについて説明する。上記実施の形態1では、読み上げ音声クラスタと自由発話音声クラスタの分類を読み上げ音声データ1-1のトークンか自由発話音声データ1-2のトークンかで単純に分類していた。これに対して、この実施の形態2による音声パターンモデル学習装置では、読み上げ音声モデルと自由発話音声モデルとの尤度の比較に基づいて各トークンをクラスタリングする。これにより、読み上げ音声の音響的特徴を持つトークンと自由発話音声データの音響的特徴を持つトークンとのクラス分類がより正確になる。従って、両クラスタの音響的特徴をより均等に反映した HMM を学習することができる。

【0072】以上のように、この実施の形態2によれ

ば、異なる発話様式（読み上げ音声、自由発話音声）で入力された音声データの音響的特徴を表す読み上げ音声データ1-1、自由発話音声データ1-2から抽出した学習データ2の発話様式ごとに応する音声パターンモデルのパラメータを求め、これら発話様式ごとの音声パターンモデルのパラメータに対応する発話様式別の音声パターンモデルを用いて学習データ2をクラスタリングして各学習データ2が属する発話様式のクラスタを決定し、クラスタリングされた各発話様式のクラスタに属する学習データのデータ量の逆数 $1/CR_c$ 、 $1/CS_c$ に、これらの総和が1となるように正規化したもの（各逆数 $1/CR_c$ 、 $1/CS_c$ に $CR_c \cdot CS_c / (CR_c + CS_c)$ ）を乗算する、（10）、（11）式参照）を、学習データ2に対する発話様式のクラスタごとのデータ量に応じたクラスタ重み係数 WR_c 、 WS_c として算出し、これらクラスタ重み係数 WR_c 、 WS_c を用いて学習データ2の各発話様式のクラスタ間におけるデータ量の違いを補正しながら、学習データ2に対応する音声パターンモデルのパラメータを求めるので、読み上げ音声モデルと自由発話音声モデルとの尤度の比較に基づいて各学習データ2をクラスタリングすることから、上記実施の形態1による構成と比較して読み上げ音声の音響的特徴を持つ学習データ2と自由発話音声データの音響的特徴を持つ学習データ2とのクラス分類がより正確になる。これにより、両クラスタの音響的特徴を均等に反映した音声パターンモデルのパラメータを得ることができるという効果が得られる。

【0073】また、この実施の形態2によれば、音声パターンモデルが隠れマルコフモデルであり、学習データ2から算出した遷移回数期待値 $\gamma^{(k)}$ （ i, j, t ）にクラスタ重み係数 WR_c 、 WS_c を乗じて、学習データ2の各発話様式のクラスタ間におけるデータ量の違いを補正したクラスタ重み付き遷移回数期待値 $\gamma_{w_c}^{(k)}$ （ i, j, t ）とし、このクラスタ重み付き遷移回数期待値 $\gamma_{w_c}^{(k)}$ （ i, j, t ）を用いてHMMのパラメータ遷移確率 $a_{i,j}^{w_c}$ 、平均値 $m_{i,j}^{w_c}$ 、分散 $v_{i,j}^{w_c}$ を求めるので、読み上げ音声の音響的特徴を持つ学習データ2と自由発話音声データの音響的特徴を持つ学習データ2とのクラス分類精度を向上させることができる。これにより、両クラスタの音響的特徴を均等に反映した音声パターンモデルのパラメータを得ることができるという効果が得られる。

【0074】なお、上記実施の形態2では、HMMで音素をモデル化する場合を説明したが、音節など他の音響単位でも構わない。さらに、上記実施の形態1、2では、音声パターンモデルとして、隠れマルコフモデルを使用する例を示したが、本願発明はこれに限らず、他の音声パターンモデルに適用することも可能である。

【0075】また、上記実施の形態では、異なる発話様式として読み上げ音声と自由発話音声について示した

が、本願発明はこれ以外の発話様式による学習データも扱うことができる。このときも、本願発明の基本概念の1つである、発話様式や発話様式のクラスタごとにおける各データ量の逆数に、これらの総和が1となるように正規化したものを重み係数、クラスタ重み係数として算出すればよい。具体的には、例えば3つの異なる発話様式の学習データに対して音声パターンモデルの学習を行う場合、発話様式ごとの学習データ量を C_1 、 C_2 、 C_3 とし、これらに対応する重み係数をそれぞれ W_1 、 W_2 、 W_3 とすると、発話様式ごとの学習データ量の逆数を $1/C_1$ 、 $1/C_2$ 、 $1/C_3$ に、これらの総和が1となるように、 $C_1 C_2 C_3 / (C_1 C_2 + C_2 C_3 + C_1 C_3)$ を乗算したものが W_1 、 W_2 、 W_3 となる。

【0076】

【発明の効果】以上のように、この発明によれば、異なる発話様式で入力された音声データの音響的特徴を表す複数種類の学習データにおける発話様式ごとの各データ量の逆数にこれらの総和が1となるように正規化したものと、学習データに対する発話様式ごとのデータ量に応じた重み係数として算出し、この重み係数を用いて学習データの各発話様式間におけるデータ量の違いを補正しながら、学習データに対応する音声パターンモデルのパラメータを求めるので、異なる発話様式に属する学習データのデータ量の不均衡を補正し、各発話様式の音響的特徴を均等に反映した音声パターンモデルを学習することができ、各発話様式に対してロバストな音声パターンモデルを得ることができるという効果が得られる。

【0077】また、文献2による従来技術と異なり、異なる発話様式で別々に音声パターンモデルの学習をすることがないので、音声パターンモデルの数を増加させない。これにより、文献2による従来技術と比較して高性能なハードウェア資源を要することがなく、コスト的にも有利である。

【0078】この発明によれば、音声パターンモデルが隠れマルコフモデルであり、重み係数を学習データから算出した遷移回数期待値に乘じて、学習データの各発話様式間におけるデータ量の違いを補正した重み付き遷移回数期待値とし、この重み付き遷移回数期待値を用いて隠れマルコフモデルのパラメータを求めるので、異なる発話様式に属する学習データのデータ量の不均衡を補正し、各発話様式の音響的特徴を均等に反映した隠れマルコフモデルを学習することができ、読み上げ音声と自由発話音声との両方に対してロバストな隠れマルコフモデルを得ることができるという効果がある。

【0079】この発明によれば、異なる発話様式で入力された音声データの音響的特徴を表す複数種類の学習データの発話様式ごとに応する音声パターンモデルのパラメータを求め、これら発話様式ごとの音声パターンモデルのパラメータに対応する発話様式別音声パターンモデルを用いて学習データをクラスタリングして、各学習

データが属する発話様式のクラスタを決定し、クラスタリングした各発話様式のクラスタに属する学習データのデータ量の逆数に、これらの総和が1となるように正規化したものを、学習データに対する発話様式のクラスタごとのデータ量に応じたクラスタ重み係数として算出し、これらクラスタ重み係数を用いて学習データの各発話様式のクラスタ間におけるデータ量の違いを補正しながら、学習データに対応する音声パターンモデルのパラメータを求めるので、各発話様式別音声パターンモデルの尤度の比較に基づいて各学習データをクラスタリングすることから、上記段落0076による構成と比較して、学習データの各発話様式の音響的特徴に対するクラス分類をより正確に行うことができ、各発話様式に対応するクラスタの音響的特徴を均等に反映した音声パターンモデルのパラメータを得ることができるという効果がある。

【0080】この発明によれば、音声パターンモデルが隠れマルコフモデルであり、クラスタ重み係数を学習データから算出した遷移回数期待値に乘じて、学習データの各発話様式のクラスタ間におけるデータ量の違いを補正したクラスタ重み付き遷移回数期待値とし、このクラスタ重み付き遷移回数期待値を用いて隠れマルコフモデルのパラメータを求めるので、学習データの各発話様式の音響的特徴に対するクラス分類精度を向上させることができ、各発話様式に対応するクラスタの音響的特徴を均等に反映した音声パターンモデルのパラメータを得ることができるという効果がある。

【図面の簡単な説明】

【図1】 この発明の実施の形態1による音声パターンモデル学習装置の構成を示すブロック図である。

【図2】 実施の形態1による音声パターンモデル学習装置における重み計算結果メモリに保持された重み係数を示す図である。

【図3】 この発明の実施の形態2による音声パターン

モデル学習装置の構成を示すブロック図である。

【図4】 実施の形態2による音声パターンモデル学習装置のクラスタリング結果メモリの内容を例示的に示す図である。

【図5】 実施の形態2による音声パターンモデル学習装置のクラスタ重み計算結果メモリに保持された内容を示す図である。

【図6】 HMMのトロポジを示す図である。

【図7】 従来の音声パターンモデル学習装置の構成を示すブロック図である。

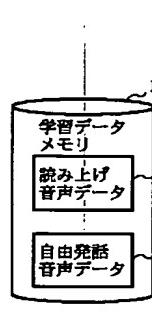
【図8】 音素テーブルの一例を示す図である。

【図9】 読み上げ音声トークンテーブルを示す図である。

【図10】 自由発話音声データトークンテーブルを示す図である。

【符号の説明】

- 1 学習データメモリ（学習データ記憶手段）、1-1 読み上げ音声データ（学習データ）、1-2 自由発話音声データ（学習データ）、2 学習データ、3 学習手段（発話様式別音声パターンモデル学習手段）、6 重み計算手段、7 重み係数、8 重み計算結果メモリ、9 重み付き学習手段、10 HMMのパラメータ（音声パターンモデルのパラメータ）、11 重み付きモデルメモリ、12 HMMパラメータ情報（発話様式別音声パターンモデルのパラメータ）、13 読み上げ音声モデルメモリ、14 HMMパラメータ情報（発話様式別音声パターンモデルのパラメータ）、15 自由発話音声モデルメモリ、16 クラスタリング手段、17 クラスタリング結果、18 クラスタリング結果メモリ、19 クラスタ重み計算手段、20 クラスタ重み計算結果、21 クラスタ重み計算結果メモリ、22 クラスタ重み付き学習手段、23 HMMパラメータ情報（音声パターンモデルのパラメータ）、24 クラスタ重み付きモデルメモリ。



【図1】

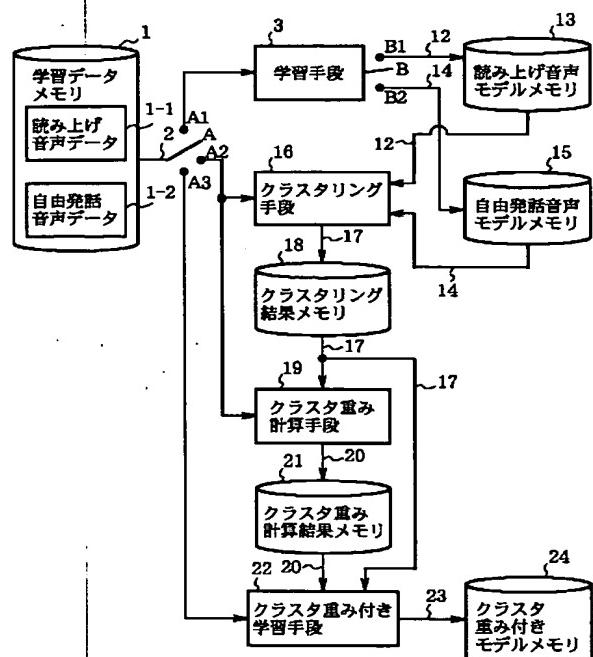
番号	音素名	重みWR (読み上げ音声データ)	重みWS (自由発話音声データ)
1	a	0.25	0.75
2	i	0.5	0.5
3	u	0.4	0.6
...
25	z	0.7	0.3

【図2】

番号	音素名
1	a
2	i
3	u
...	...
25	z

【図8】

【図3】



【図4】

トークン番号	音素表記	クラスタリング結果
1	a	R
2	a	R
...
120	a	S
121	i	S
122	i	R
...
300	i	R
...
915	z	R
916	z	S
...
1020	z	S
1021	a	S
1022	a	S
...
1060	a	R
1061	i	S
1062	i	S
...
1100	i	S
...
1521	z	R
1522	z	S
...
1580	z	S

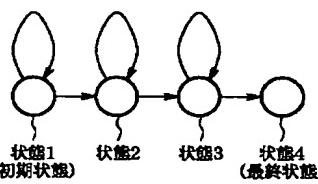
【図9】

トークン番号	音素表記
1	a
2	a
...	...
120	a
121	i
122	i
...	...
300	i
...	...
915	z
916	z
...	...
1020	z

【図5】

番号	音素名	重みWRc (読み上げクラスタ)	重みWSC (自由発話クラスタ)
1	a	0.2	0.8
2	i	0.5	0.5
3	u	0.4	0.6
...
25	z	0.7	0.3

【図6】



トークン番号	音素表記
1021	a
1022	a
...	...
1060	a
1061	i
1062	i
...	...
1100	i
...	...
1521	z
1522	z
...	...
1680	z

【図7】

